

6-1-1995

IMAGE REPRESENTATION AND COMPRESSION USING MULTISPLINES AND ADAPTIVE SCALE-SPACE ATOM SELECTION

Shankar Moni

Purdue University School of Electrical and Computer Engineering

R. L. Kashyap

Purdue University School of Electrical and Computer Engineering

Follow this and additional works at: <http://docs.lib.purdue.edu/ecetr>

Moni, Shankar and Kashyap, R. L., "IMAGE REPRESENTATION AND COMPRESSION USING MULTISPLINES AND ADAPTIVE SCALE-SPACE ATOM SELECTION " (1995). *ECE Technical Reports*. Paper 139.
<http://docs.lib.purdue.edu/ecetr/139>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

IMAGE REPRESENTATION AND
COMPRESSION USING MULTISPLINES
AND ADAPTIVE SCALE-SPACE
ATOM SELECTION

SHANKAR MONI
R. L. KASHYAP

TR-ECE 95-18
JUNE 1995



SCHOOL OF ELECTRICAL
AND COMPUTER ENGINEERING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907-1285

IMAGE REPRESENTATION AND COMPRESSION USING MULTISPLINES
AND ADAPTIVE SCALE-SPACE ATOM SELECTION*

Shankar Moni and R. L. Kashyap

School of Electrical Engineering
1285 Electrical Engineering Building
Purdue University
West Lafayette, IN 47907-1285

*THIS WORK WAS SUPPORTED IN PART BY THE OFFICE OF NAVAL RESEARCH UNDER CONTRACT ONR N00014-91-J-4126.

ABSTRACT

We define a new stochastic process for representing images. We call this process the Ordered-Tree process (OTP). We show the existence of such a process and derive the optimal compression algorithm for such a process. Experimental results have indicated that the algorithm outperforms many existing image compression algorithms.

In order to define the stochastic process, we first define a Tree-Structured analysis (TSA). This is a generalization of a multiresolution analysis (MRA) that extracts only those properties of an MRA that serve well in image compression. In particular, we place no requirement on self-similarity or orthogonality of basis functions.

We give a detailed example of the TSA and the OTP. Several theorems are proved that explore the properties of the TSA and the OTP.

Keywords: Multiresolution, Multisplines, Wavelets, Tree-Structured Analysis, Ordered-Tree Stochastic Process.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
LIST OF FIGURES	v
1. INTRODUCTION	1
2. SUMMARY OF RESULTS	4
3. TREE-STRUCTURED ANALYSIS	5
3.1 Definitions	5
3.2 Multisplines: An example of Tree-Structured Analysis	8
3.2.1 Construction of Multisplines	8
3.2.2 Multisplines form an Tree-Structured Process	9
4. IMAGE REPRESENTATION USING THE ORDERED-TREE STOCHAS- TIC PROCESS	12
4.1 Definition of the Ordered-Tree Process	12
4.2 Optimal Compression Algorithm for an Ordered-Tree Process	13
5. PROPERTIES AND EXTENSIONS OF THE ORDERED-TREE PROCESS	16
5.1 Properties	16
5.2 A Stochastic Process with Decreasing Detail-Energy	17
6. EXPERIMENTAL RESULTS	19
6.1 Compression	19
6.2 Noise Removal	19
6.3 Other Properties	22
7. CONCLUSIONS	24
LIST OF REFERENCES	25

	Page
APPENDIX PROOFS OF THEOREMS	27

LIST OF FIGURES

Figure	Page
3.1 Illustration of the tree T. Some nodes are numbered. Note the one to one correspondence between nodes on the tree and dyadic rationals in $(0, 1]$.	7
4.1 Flowchart of algorithm. The second and third steps of the algorithm are repeated until the desired rate or distortion level is achieved.	15
6.1 Comparison of degradation in image quality as number of coefficients is reduced. The Multispline approach outperforms both Haar-wavelet and DCT methods at both high and low compression ratios. Instead of the Haar wavelet, other orthogonal wavelets may be used, but these are also jagged in appearance.	21
6.2 Gaussian Noise: Comparison of our restoration method to the random field approach. In figure (b), blocking can be observed. In figure (c), there is no blocking but noise is not removed. The multispline approach (figure (d)) has no blocking artifacts, removes noise, and additionally performs compression.	23

1. INTRODUCTION

Information in images is present at various scales and various spatial locations. It is therefore natural to represent an image using a tree where the root corresponds to information at the coarsest scale and the branches represent information at finer scales. In 1983, Burt and Adelson [2] proposed a tree-structured algorithm that enabled computing the approximation of a discrete-time signal (or an image) at any particular scale using only the information at the previous scale. A follow-up to this was Mallat's continuous-time multiresolution analysis [9] which uses wavelets to decompose a function. Wavelet-based algorithms have generated a lot of interest in the image processing community because there is a need for techniques that are capable of zooming in on details wherever present and retaining coarse resolution where details are absent.

In order to compress an image, we would like to utilize the fact that details are usually present only in certain spatial locations of the image. The wavelet decomposition computes the wavelet coefficients (and thereby estimates the amount of detail present) at *all* spatial locations of the image. Wavelet based image compression schemes first compute the entire wavelet transform and then postprocess the results in order to achieve compression [7] [10]. This requires a significant overhead in terms of computational and memory resources. What we would ideally like to have is a method that tells us up-front in *which part* of the image wavelets are needed. Some attempts have been made to integrate wavelets into a stochastic framework [1], [4], [17] but these are not designed to pinpoint the locations where wavelets are required.

In this paper, the objective is image compression. We assume that a noiseless image is available to us. We index basis functions on a tree and construct a stochastic process from it. The image data together with the stochastic assumptions give us an idea of *where* (i.e. in which part of the image) to add more detail. The approximation that results from using these stochastic assumptions results in compression that captures visually important details. Note that simply thresholding wavelet coefficients may not result in details that are visually important. For example, a dot near the corner of an image may yield a high wavelet coefficient but may be visually insignificant. We propose an approach that performs compression by jointly considering the spatial information, the magnitude of the coefficients, and the stochastic assumptions.

We start by constructing a tree on which to index the basis functions. We then associate some basis functions with each node of the tree. We require that any

functions associated with any finite collections of nodes be linearly independent. We also require that as we consider nodes deeper down the tree, the support of the functions associated with the node decrease exponentially with the depth of the node. Our approach is to extend the ideas of a multiresolution analysis by extracting the essential features that are useful in analysing images. We call this approach a "Tree-Structured Analysis".

We then construct a stochastic process using the Tree-Structured analysis. We order the nodes of the tree in such a manner that for any n , the first n nodes form a subtree. For the basis functions associated with the n th node, we assign coefficients (weights) that decrease with n . The ordering of the tree and the values of the coefficients are both random. The stochastic process is constructed by summing up the weighted basis functions. In this approach, some regions of the image get more detail than others. The locations of high detail and the amount of detail are both random. We call this process the "Ordered-Tree Process".

The optimal algorithm to compress such a process is then developed. The algorithm is simple and seeks the subtree that grabs the maximum energy from the given data. As shown in the experimental results, the algorithm performs well on real images. An encouraging property of the algorithm that we prove is that it zooms in on discontinuities.

The approach outlined above is in contrast to the multiresolution analysis in many ways. Firstly, the multiresolution analysis uses a single function at each node whereas we use more than one (typically two). Secondly, the basis functions of a multiresolution analysis are translates and dilates of a single "mother" function, whereas we place no such restriction. Third, the multiresolution analysis requires that the basis functions form a Riesz basis whereas we only require linear independence of a *finite* collection. Fourth, we place no requirement on self-similarity of the basis functions.

The advantage of placing multiple functions (typically two) at each node is essentially that we get better compression. The reason for this is that we get more degrees of freedom to approximate the image where details are present. We can thus do a better job this way than having to resort to several more wavelets (which are all of the same essential shape). In real image compression applications, there is not much advantage to having a unique representation of a function in an infinite basis. This is the Riesz basis requirement that the multiresolution analysis imposes that we dispense with. Rather, we only require that any finite set of functions is linearly independent. Finally, if an image is truly self-similar in nature, it is advantageous to use basis functions that exhibit self-similarity. However, if that is not the case, it is not clear that self-similar basis functions perform a better job at approximating the image. We therefore do not require the functions to be self-similar. Note, however, that we do expect the support of the functions to behave in a manner very similar to that of compactly supported wavelets. Further, we do require that any finite collection of these functions be linearly independent. These two requirements are sufficient to capture most of the useful properties of a multiresolution analysis.

The other properties of a multiresolution analysis contribute mathematical elegance but are not necessarily beneficial for image compression.

In the interest of clarity, we develop the theory for one-dimensional images supported on the unit interval $(0,1]$. The extension of our ideas to two dimensions is straightforward. We summarise our results in chapter 2. The Tree-Structured analysis is introduced in chapter 3 and some of its properties are explored. The Ordered-Tree process and optimal compression algorithm are presented in chapter 4. Properties and extensions of the process are presented in chapter 5. The conclusions are in chapter 6, and this is followed by an appendix in which proofs are included.

2. SUMMARY OF RESULTS

We define a Tree-Structured Analysis as a generalization of a Multiresolution Analysis. We lay down the requirements of a Tree-Structured Analysis and give a detailed example. The example is to use multiple spline-based, multiresolution analyses simultaneously.

Let $\phi(x)$ denote a scaling function in a multiresolution analysis. We show that using only the even translates of all dilates of ϕ results in the same space as using all translates of a fixed dilate of ϕ . The latter is the multiresolution approach. Our approach is to start with the former set of functions but choose them optimally under some stochastic assumptions thereby gaining more compression.

We define an Ordered-Tree stochastic process. We prove the existence of such a process. We show that in any interval $(x - \epsilon, x + \epsilon) \subset (0, 1]$ details are added infinitely often with probability one. Note, however, that the details may not be large enough in magnitude to be visually significant. We show that we can construct the Ordered Tree process using multisplines so that the resulting stochastic process has M continuous derivatives almost everywhere. We derive an expression for the estimate of the correlation between two points using the compressed approximation of the image.

We derive the optimal compression algorithm for an Ordered Tree process. We examine some properties of this algorithm. In particular, we find that the algorithm zooms in on discontinuities.

3. TREE-STRUCTURED ANALYSIS

A multiresolution analysis (MRA) [9] has the ability to represent a function at varying levels of detail. It is well known that some spatial regions of an image may have more details than others. The drawback of the MRA is that it does not adapt the choice of basis functions to the given data. Rather, it simply assigns the same level of detail at all spatial locations.

In this chapter, we extract the essential properties of the MRA that are useful in the analysis of nonstationary signals. We then enhance these properties and formulate what we call a "Tree-Structured Analysis" (TSA). We give some examples of a TSA and explore some of the similarities and differences between a TSA and an MRA. Specifically, we show that a TSA has all the multiresolution properties of an MRA *without* the use of wavelets.

3.1 Definitions

Throughout the development, our region of interest is the unit interval $(0, 1]$ whose indicator function will be denoted by $\chi_{(0,1]}$. Let $D_{(0,1]}$ be the set of all dyadic rationals in $(0, 1]$. If $d \in D_{(0,1]}$, define $\gamma(d)$ and $\kappa(d)$ by the equations

$$\begin{aligned}\kappa(d) &= \min_{j \geq 0} \{2^j d : 2^j d \in \mathbf{Z}\} \\ \gamma(d) &= \arg \min_{j \geq 0} \{2^j d : 2^j d \in \mathbf{Z}\}\end{aligned}\tag{3.1}$$

Note that $\kappa(d)$ and $\gamma(d)$ are both integers. Further, we have

$$d = \frac{\kappa(d)}{2^{\gamma(d)}} \text{ with } \kappa(d) \text{ odd.}$$

We refer to the elements d of $D_{(0,1]}$ as scale-space atoms. Also, if $D \subset D_{(0,1]}$, we will call D a scale-space set. Some examples of scale-space atoms are $d = 1$, $d = \frac{1}{2}$, $d = \frac{3}{16}$, $d = \frac{5}{32}$, $d = \frac{17}{32}$, and $d = \frac{117}{128}$. Some examples of scale-space sets are $D = \{1, \frac{1}{2}, \frac{3}{16}\}$, $D = \{\frac{3}{16}, \frac{1}{128}, \frac{47}{64}\}$, $D = \{\frac{5}{8}, \frac{13}{16}, \frac{121}{128}, \frac{27}{32}\}$, etc.

Let T be the tree shown in figure (3.1). The root node is 1, and it has one child, $\frac{1}{2}$. This, and all other nodes have two children each. In the tree T , the nodes d at level j , are successively numbered

$$\frac{\kappa(d)}{2^{\gamma(d)}} \text{ where } \kappa(d) = 1, 3, 5, \dots, 2^{\gamma(d)} - 1.$$

The numbering convention is shown in that figure. It is clear that there is a one-to-one correspondence between dyadic rationals in $D_{(0,1]}$ and the nodes of the tree T . Hereafter, we will make no distinction between nodes of T and elements of $D_{(0,1]}$, and we will use the terms "node" and "scale-space atom" interchangeably.

Let M and K be fixed integers with $-1 \leq M < K - 1$. In our examples, we will use piecewise polynomials with M continuous derivatives composed of polynomials of degree at most $K - 1$. However, in the general case, the only significance of M and K is that we can have $K - 1 - M$ basis functions at each node. This is in contrast to the multiresolution analysis which has just one basis function at each node.

Notations:

1. Let \mathbf{Z} denote the set of integers and $\mathbf{Z}_{M,K} = \{m \in \mathbf{Z} : M + 1 \leq m \leq K - 1\}$. Let \mathbf{N} denote the set of natural numbers.
2. For any scale-space set D , the number of nodes in D will be denoted by $|D|$. A scale-space set D is said to be finite if $|D| < \infty$.
3. The set of square integrable functions on $(0, 1]$ will be denoted by $\mathcal{L}_{(0,1]}^2$.

Definition: A Tree-Structured Analysis is a set of bounded functions ϕ_d^m , $m \in \mathbf{Z}_{M,K}$, $d \in D_{(0,1]}$, which satisfies the following properties.

- P1: The collection of functions $\{\phi_d^m : m \in \mathbf{Z}_{M,K}, d \in D\}$ is dense in $\mathcal{L}_{(0,1]}^2$.
- P2: For any finite scale-space D , the collection of functions $\{\phi_d^m : m \in \mathbf{Z}_{M,K}, d \in D\}$ forms a linearly independent set.
- P3: There exists a nonnegative constant C (which may depend on M and K) so that the support of the functions ϕ_d^m is contained in the interval $(\frac{\kappa(d)-C}{2^{\gamma(d)}}, \frac{\kappa(d)+C}{2^{\gamma(d)}}]$.

Remarks:

1. The multiresolution analysis is defined on $L^2(\mathbf{R})$, whereas the Tree-Structured Analysis is defined on $\mathcal{L}_{(0,1]}^2$. The reason for this is because images are compactly supported and we need to be able to zoom in on specific regions of this fixed interval. The tree structure is especially suited to zooming in on regions of a fixed interval and experimental results confirm the superiority of this method.

2. In a multiresolution analysis, the basis functions are required to form a Riesz basis [3]. Therefore, any function $f \in L^2(\mathbf{R})$ must have a unique representation in terms of the basis functions. In other words, we may write

$$f = \sum_{d \in D_{(0,1]}} c_d \phi_d$$

and the c_d 's are unique. In the Tree-Structured Analysis, we require the c_d 's to be unique for *finite* subsets D of $D_{(0,1]}$.

3. A useful property to have is localization of the basis functions. This is ensured by property P3. In contrast, the basis functions of a multiresolution analysis may or may not have this property.

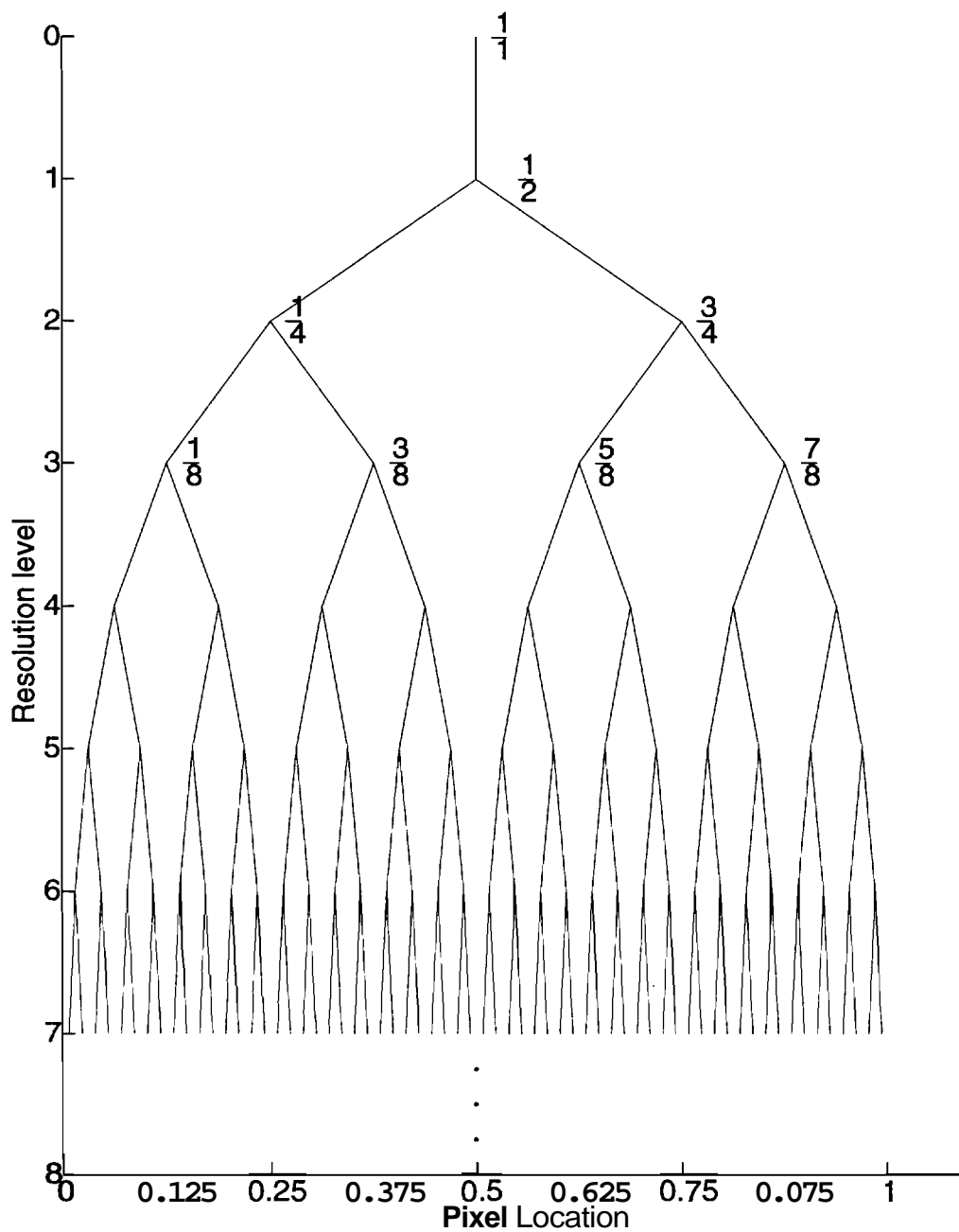


Figure 3.1 Illustration of the tree T . Some nodes are numbered. Note the one to one correspondence between nodes on the tree and dyadic rationals in $(0, 1]$.

4. In a multiresolution analysis, the basis functions are required form a shift-invariant basis, i.e. if $\phi(x)$ is a basis function, then so is $\phi(x - 1)$. Also they are required to be self-similar, i.e. there exists a sequence $\{p_k\}_{k \geq 1}$ so that

$\phi(x) = \sum_{k=-\infty}^{\infty} p_k \phi(2x - k)$. In the Tree-Structured analysis, we impose no such requirement. The main reason is that such self-similar functions may not be the best choice of basis functions for every image. Further, in continuous-time systems, such self-similar functions may not be realizable.

3.2 Multisplines: An example of Tree-Structured Analysis

Images are well modeled by piecewise polynomials. We may therefore try to represent the image using spline-based non-orthogonal wavelets [3], which are composed of piecewise polynomials. However, the difference between the polynomial order (of the piecewise polynomials) and the number of continuous derivatives (of a function composed of these wavelets) is fixed. For example, a function composed of 3rd order polynomials come with 2 continuous derivatives, which usually results in excessive overshoots in the fitted curve [6]. For image representation, we may like to have a function with a small number of continuous derivatives (say 0) composed of piecewise polynomials with large order (say 3). One possibility is to use multiple spline-based wavelets simultaneously because that gives us control on both the polynomial order and the number of continuous derivatives. However, in that case, the usual wavelet decomposition is no longer valid (i.e., we cannot use the dual wavelet to compute the wavelet coefficients). We therefore need an appropriate family of piecewise polynomial functions together with an inexpensive algorithm for computing the coefficients.

We are interested in functions with M continuous derivatives. The functions are to be composed of piecewise polynomials of degree at most $K - 1$. We define a space of splines $\mathcal{S}_D^{K,M}$ that satisfies these conditions. The subscript D will denote a set of scale-space atoms that specifies which translates and dilates of the splines will be used. Since splines themselves are piecewise polynomial, D indirectly specifies where the knots of the piecewise polynomials will lie. We adopt the convention that $M = 0$ denotes continuous functions and $M = -1$ denotes functions that are possibly discontinuous.

3.2.1 Construction of Multisplines

The zeroth order spline is defined as

$$\phi^0(x) = \chi_{(0,1]}(x) \quad [3.2]$$

and the m th order spline is defined recursively by the relation

$$\phi^m(x) = [\phi^{m-1} * \phi^0](x) \quad [3.3]$$

where $*$ denotes the convolution operator. By indexing the splines using $d \in D_{(0,1]}$, we can denote the dilates and even translates of splines as follows.

$$\phi_d^m = \phi^m(2^{\gamma(d)}x - \kappa(d) + 1)\chi_{(0,1]} / \|\phi^m(2^{\gamma(d)}x - \kappa(d) + 1)\chi_{(0,1]}\| \quad [3.4]$$

where $\|\cdot\|$ denotes the \mathcal{L}^2 norm. Note that since $\kappa(d)$ is odd, we must have $\kappa(d) - 1$ even. This ensures that equation (3.4) indexes only even translates of the splines. Also note that we have multiplied the function ϕ^m by the indicator function of the unit interval. In other words, we truncate those parts of the function that reside outside the unit interval. Further, the denominator in equation (3.4) ensures that ϕ_d^m is a unit energy function. Let

$$\mathcal{S}_D^{K,M} = \{ f \in \mathcal{L}_{(0,1]}^2 : f = \sum_{m \in \mathbf{Z}_{M,K}} \sum_{d \in D} c_d^m \phi_d^m \}. \quad [3.5]$$

Then $\mathcal{S}_D^{K,M}$ is a Hilbert space when endowed with the norm

$$\langle f, g \rangle = \int_0^1 f(x)g(x) dx \quad f, g \in \mathcal{S}_D^{K,M} \quad [3.6]$$

We refer to the set $\mathcal{S}_{D_{(0,1]}}^{K,M}$ as the space of multisplines with M continuous derivatives composed of $(K-1)$ th order polynomials. It is clear that for any $D \subset D_{(0,1]}$, we have $\mathcal{S}_D^{K,M} \subset \mathcal{S}_{D_{(0,1]}}^{K,M}$. The ability to choose D adaptively is the one of the most important points in this research. The algorithm that chooses D makes use of the tree-based indexing scheme presented offered by the Structured-Tree Analysis.

3.2.2 Multisplines form an Tree-Structured Process

We now show that multisplines form a Tree-Structured Process. Property P3 is easy to verify by setting $C = K - 1$. In order to verify properties P1 and P2, we need two theorems. We first state a definition and then the theorems.

Definition: A scale-space set D is saturated if $D = \{\frac{k}{2^j} : k = 1, \dots, 2^j\}$ for some j . When necessary, we will refer to this D as the saturated scale-space set at resolution

Theorem 1 Let D be any finite scale-space set. Let ϕ_d^m be as defined in equation (3.4). Then the set of functions $\{\phi_d^m : d \in D, m \in \mathbf{Z}_{M,K}\}$ form a linearly independent set of functions.

Proof: The proof is in the appendix. ■

Theorem 2 Let D_j be the saturated scale-space set at resolution j . Then

$$\begin{aligned} \text{span}\{\phi_d^m : d \in D_j, m \in \mathbf{Z}_{M,K}\} = \\ \text{span}\{\phi^m(2^j x - k)\chi_{(0,1]} : k = 0, 1, \dots, 2^j - 1, m \in \mathbf{Z}_{M,K}\} \end{aligned}$$



Proof: The proof is in the appendix. ■

We now consider the relationship between multisplines and piecewise polynomials. Given a scale-space set D , define

$$\begin{aligned} \mathcal{S}^{0;M} &= \left\{ f \in \mathcal{L}_{[0,1]}^2 : f = \sum_{m=0}^M c_{m,1} \phi_1^m \right\} \\ \tilde{\mathcal{S}}_D^{K,M} &= \mathcal{S}^{0;M} \cup \mathcal{S}_D^{K,M} \\ \mathcal{P}_D^{K,M} &= \left\{ \mathbf{f} \in \mathcal{L}_{[0,1]}^2 : \begin{array}{l} \mathbf{f} \text{ is composed of piecewise polynomials} \\ \text{of degree at most } K - 1 \\ \text{with knots at the elements of } D \\ \text{and } \mathbf{f} \text{ has } M \text{ continuous derivatives} \end{array} \right\} \end{aligned}$$

Theorem 3 Let M, K be integers with $-1 \leq M < K$. Let D be a saturated scale space set. Then

$$\tilde{\mathcal{S}}_D^{K,M} = \mathcal{P}_D^{K,M}$$

where $\mathcal{P}_D^{K,M}$ is the space of piecewise polynomials functions which are composed of piecewise polynomials of degree at most $K - 1$ with knots at the elements of D , and possess M continuous derivatives.

Proof: The proof is in the appendix. ■

Remarks:

1. The content of Theorem 2 is as follows. Let V_j^m be the (conventional) multiresolution space based on the spline of order m as given in [3]. Then the *even* translates (of all dilates upto j) of the splines ϕ^m forms a basis for the direct sum of the spaces V_j^m , $m \in \mathbf{Z}_{M,K}$. Thus, we do not need wavelets to construct a multiresolution. Using only the *even* translates of the scaling function will do the job.

2. If we set $m = K - 1 = M + 1$, then we are dealing with a single spline of order m . In this case, it has already been shown that this spline can be used as a scaling function to generate a multiresolution analysis [3]. Further, it has been shown that this set of functions is dense in $L^2(\mathbf{R})$. It is therefore dense in $\mathcal{L}_{[0,1]}^2$. Further, adding more functions by increasing $K - M$ will not alter this property. Thus, property P1 follows from Theorem 2.

3. Property **P2** follows immediately from Theorem 1.

4. Of late, there has been some interest in using multiple scaling functions to construct multiwavelets [16]. The motivation for constructing multiwavelets is to be able to construct new orthogonal wavelets. However, the wavelets so constructed are very jagged in appearance and not particularly suited for image processing applications. On the other hand, our multisplines use the translates and dilates of multiple splines. These are *not* orthogonal functions. In our approach, we relax the requirement of orthogonality and focus instead on functions that can *economically* represent an image.

5. Theorem 3 says that if we have a saturated scale-space set D , then we are essentially dealing with the space of piecewise polynomials with knots at dyadic rationals. This theorem provides the reassurance that if we limit ourselves to a large but finite resolution, then we do indeed eventually get the desired space of piecewise polynomials.

4. IMAGE REPRESENTATION USING THE ORDERED-TREE STOCHASTIC PROCESS

We would like to construct a stochastic process that is effective at representing an image. Since the image data is usually available to us, the emphasis is on compression of data rather than prediction. Any image can be represented in the form

$$g = \sum_{d \in D_{(0,1]}} \sum_{m \in \mathbb{Z}_{M,K}} c_d^m \phi_d^m(x) \quad [4.1]$$

where the coefficients c_d^m determine the image. Recall that we can establish a one to one correspondence between the atoms in $D_{(0,1]}$ and the nodes of the tree T . Let the nodes of the tree T be ordered so that for each n , the first n nodes form a subtree. In our stochastic process, this ordering is random. The n th coefficient $c_{d_n}^m$ is chosen randomly from a probability distribution that depends on the ordering of T as well as on the previously chosen coefficients.

We start this chapter by defining the Ordered-Tree Process, then show the existence of such a process and finally explore some of its properties. This stochastic process has some interesting and important properties such as the following. In any interval $(x - \epsilon, x + \epsilon)$, details are added infinitely often with probability 1. Other properties are also discussed.

4.1 Definition of the Ordered-Tree Process

Definition: The parent of a node d is the node $(\lfloor \kappa(d)/2 \rfloor) / 2^{\gamma(d)}$. The children of a node d are the nodes $(2\kappa(d) - 1) / 2^{\gamma(d)+1}$ and $(2\kappa(d) + 1) / 2^{\gamma(d)+1}$.

Definition: A subtree is a set B of nodes containing 1 and some or all other nodes, satisfying the condition that the parent of every node (other than 1) in B is also in B . The set of children of a subtree B , denoted by $\text{chil}(B)$, is the set of all nodes which are children of nodes in B but are not themselves in B .

Let the nodes of the tree T be ordered as $\{d_1, d_2, \dots\}$ where the ordering satisfies the condition that for any n , the first n nodes form a subtree. That is, for any n , $B_n = \{d_1, d_2, \dots, d_n\}$ is a subtree. It is easy to see that given B_{n-1} , there are n possible choices for d_n that satisfy the requirement that B_n be a subtree. The node d_n may be chosen as one of the children of B_{n-1} . We set the conditional distribution to be uniform, i.e.

$$P(d_n = \check{d} / B_{n-1}) = \begin{cases} \frac{1}{n}, & \check{d} \in \text{chil}(B_{n-1}) \\ 0 & \text{otherwise} \end{cases} \quad [4.2]$$

This specifies the distribution from which nodes are chosen.

It remains to specify how the coefficients c_d^m are selected. In order to do this we use the following notation. Let D be any scale-space set and let

$$\mathcal{S}_D = \{ \phi_d^m : d \in D, m \in \mathbf{Z}_{M,K} \}$$

Let g be any function and denote the projection of g onto \mathcal{S}_D as $S_D g$. The energy in this projection is

$$\|S_D g\|^2 = \langle S_D g, S_D g \rangle$$

We can now define the Ordered-Tree process.

Definition: An Ordered-Tree process is

$$g(x) = \sum_{i=1}^{\infty} \sum_{m \in \mathbf{Z}_{M,K}} c_{d_i}^m \phi_{d_i}^m(x) \quad [4.3]$$

where the d_i 's are chosen as in equation (4.2) and the $c_{d_i}^m$'s satisfy the following condition:

$$\|S_{\check{D}} g\|^2 > \|S_{D_n} g\|^2 \quad [4.4]$$

where \check{D} is any scale-space set satisfying $|\check{D}| = n$ and $\check{D} \neq B_n$.

Theorem 4 There exists an Ordered-Tree process $g(x)$ as in equation (4.3) satisfying equation (4.4).

Proof: The proof is included in the appendix. ■

4.2 Optimal Compression Algorithm for an Ordered-Tree Process

In this section, we derive the optimal compression algorithm for the Ordered-Tree process. We then explore the possibility of using this algorithm on piecewise constant functions with a finite number of jumps. Interestingly, we find that the tree trickles down to the locations of the jumps. It is not clear at present whether the piecewise constant function can nontrivially be represented as an instantiation of the Ordered-Tree process. However, it is encouraging to know that the algorithm performs well on such functions.

Let D, \check{D} be scale-space sets, each with n nodes. Given an Ordered-Tree process

$$g(x) = \sum_{d \in D_{[0,1]}} \sum_{m \in \mathbf{Z}_{M,K}} \phi_d^m(x)$$

we consider approximations of the form

$$g_D(x) = \sum_{d \in D} \sum_{m \in \mathbf{Z}_{M,K}} \phi_d^m(x)$$

where D is a scale-space set.

Definition: Let $\{D_n\}_{n=1}^\infty$ be a sequence of scale-space sets. A sequence $\{S_{D_n}g\}_{n=1}^\infty$ of approximations of a function $g(x)$ is said to be the optimal compression sequence if for each n ,

$$\|S_{D_n}g\| > \|S_{\check{D}}g\| \quad \forall \check{D} \text{ with } |\check{D}| = |D_n|.$$

Algorithm: Let g be the Ordered-Tree process to be approximated.
 Repeat steps 2 and 3 until the desired rate or distortion is achieved. Step 1. Start at the root of the tree, i.e. set $B = \{1\}$.
 Step 2. Our current approximation of g is $S_B g$.
 Step 3. Of all children of B , find the node d that maximizes $\|S_{B \cup \{d\}}g\|$. Set $B = B \cup \{d\}$.
 Repeat steps 2 and 3 until the desired rate or distortion is achieved. This is shown in figure (4.1).

Theorem 5 For an Ordered-Tree process, the algorithm in figure (4.1) is optimal.

Proof: The proof is in the appendix. ■

We now consider what happens if the algorithm is used on a piecewise constant function. In particular, we would like to know if the subtree that results from the algorithm zooms in on the discontinuities of such a function.

Theorem 6 Let $0 < x_1 < x_2 < \dots < x_N = 1$ and let $p(x) \sum_{i=1}^N c_i \chi_{(x_{i-1}, x_i]}$ be a piecewise constant function on $(0, 1]$. Let $\epsilon > 0$ be a fixed small number.

1. If the algorithm is used with multisplines with $M = -1, K = 1$ (i.e., only the zeroth order spline), then the algorithm chooses nodes in the interval $(x - \epsilon, x + \epsilon)$ infinitely often if and only if x is a discontinuity point (i.e. $x \in \{x_1, \dots, x_{N-1}\}$).
2. If the algorithm is used with multisplines with $M > -1$ (i.e., multiple splines starting with first order), then the algorithm chooses nodes in the interval $(x - \epsilon, x + \epsilon)$ infinitely often if x is a discontinuity point (i.e. $x \in \{x_1, \dots, x_{N-1}\}$).

Proof: The proof is in the appendix. ■

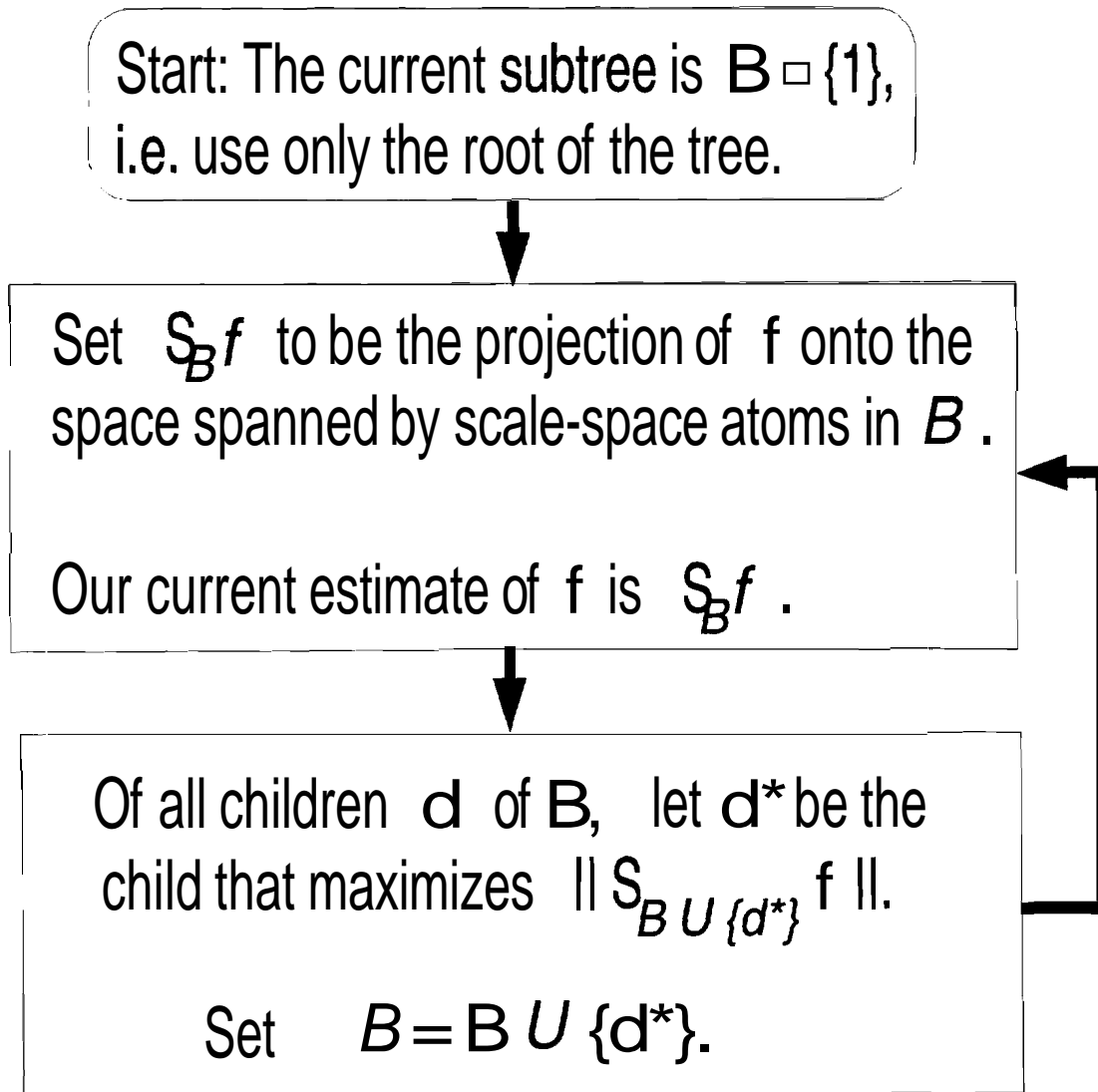


Figure 4.1 Flowchart of algorithm. The second and third steps of the algorithm are repeated until the desired rate or distortion level is achieved.

5. PROPERTIES AND EXTENSIONS OF THE ORDERED-TREE PROCESS

In this chapter, we first outline some properties of the Ordered-Tree process. Then, we look at an extension of this process.

5.1 Properties

We now outline some of the properties of the Ordered-Tree Process.

Theorem 7 Let

$$(L_d^m, \tilde{L}_d^m) = \inf\{(c, \tilde{c}) : \text{support}(\phi_d^m) \subset (\frac{\kappa(d) - c}{2^{\gamma(d)}}, \frac{\kappa(d) + \tilde{c}}{2^{\gamma(d)}}]\} \quad [5.1]$$

Suppose that for each $d \in D_{[0,1]}$, there is at least one $m \in \mathbf{Z}_{M,K}$ so that

$$L_d^m \geq 1, \tilde{L}_d^m \geq 1$$

Let $x \in (0,1)$ and let $\epsilon > 0$ be an arbitrarily small positive number. Then with probability 1, the tree visits $(x - \epsilon, x + \epsilon)$ infinitely often. In other words, for any interval $(x - \epsilon, x + \epsilon)$, details are added infinitely often.

Proof: The proof is in the appendix. ■

Corollary 1 For the Ordered Tree process based on multisplines, with probability 1, details are added infinitely often in any interval $(x - \epsilon, x + \epsilon)$.

Proof: The proof is in the appendix. ■

Theorem 8 Let $g(x)$ be the Ordered Tree process and $g_n(x)$ be the restriction of $g(x)$ to the first n nodes. Then

$$\frac{E[g_n(x)g_n(y)]}{E[g(x)g(y)]} = \frac{[\Phi_n(x)^T \Pi_b \Phi_n(y)]^2}{\Phi_n(x)^T \Pi_b \Phi_n(y) \Phi_n(x)^T \Pi_b \Phi_n(y)}$$

where $\Phi_n(x) = [\phi_{d_1}^{M+1}(x), \dots, \phi_{d_n}^{K-1}(x), \dots, \phi_{d_n}^{M+1}(x), \dots, \phi_{d_n}^{K-1}(x)]^T$.

Proof: The proof is in the appendix. ■

Remarks:

1. Theorem 4 says that details will always be added in any interval. However, this may happen for very large n . In this case, the $\mathcal{L}_{[0,1]}^2$ energy added is very small. Recall

that the energy added decreases as the iteration number n increases. We would like to reconstruct the process in a manner that retrieves the maximum energy for minimum number of coefficients. In the next chapter, we develop the algorithm that does precisely that. This is in contrast to the multiresolution algorithm that allocates coefficients uniformly over the entire image without any regard to the amount of energy involved.

2. Consider the Multispline Ordered-Tree process and let $x, y \in (0, 1]$. The correlation coefficient between x and y will be less if the tree has attained a large depth near x and more if the tree has not traversed deep near x . This is simply because the splines ϕ_d^m are smooth functions whose variation increases as $\gamma(d)$ increases.

3. If the basis functions $\phi_d^m(x)$ are continuous functions of x , then the correlation coefficient is a continuous function of its arguments.

5.2 A Stochastic Process with Decreasing Detail-Energy

In this section, we define a stochastic process that has decreasing detail-energy (DDE). This is a variation of the ordered-tree process. We also show that the algorithm in figure (4.1) is optimal for this process too, though in a different sense.

In the MRA, the detail space W_j is orthogonal to the coarse space V_j and it complements it to form the fine resolution space V_{j+1} [8]. In the DDE process, as the resolution j increases, the energy in the detail space decreases. The exact definition of the DDE process is given below.

A DDE process is a random function f that obeys the following property:

- A1: Let B be a subtree. Let $d, d' \notin B$.
If $d' \in \text{desc}(\{d\})$ then $P(C_f(d/B) > C_f(d'/B)) > 0.5$, where $P(\cdot)$ denotes the probability of an event.

In order to prove the existence of such a process, one must specifically consider which function(s) are used as scaling functions to form the multiresolution analysis. In our case, the scaling functions used are the splines d^m , $m = M+1, \dots, K-1$. In the theorem below, we show the existence of such a process for $M=0, K=1$.

Theorem 9 For the multispline family characterized by $M=0, K=1$, there exists a random function f that satisfies the property A1.

Proof: The proof is in the appendix. ■

Although we have not proved the existence of such a process for other choices of M and K , we have successfully used many values of M and K in our experiments. Our experiments find the optimal function with respect to the criterion defined below.

Definition: Let f be the function to be estimated and let F be a set of estimates of f . Let $\hat{f}, \check{f} \in F$. The Highest Probability Estimate of f in F , denoted by $\hat{f}_{HP(F)}$, is the estimate that has the highest probability of minimizing the $\mathcal{L}_{(0,1]}^2$ error, i.e.

$$\hat{f}_{HP(F)} = \arg \max_{\hat{f} \in F} P(\|\hat{f} - f\| \leq \|\check{f} - f\| \quad \forall \check{f} \in F, \check{f} \neq \hat{f}).$$

Let F_1 be the set of functions that uses only one scale-space atom, i.e. the cardinality of the scale-space set D in \mathcal{S}_D is 1. The problem of finding $\hat{f}_{HP(F_1)}$ is equivalent to answering the following question, “If we have to estimate f using only 1 scale-space atom, which atom should we choose?”. The answer to this is shown (in theorem 4 below) to be the root node. Having approximated f using a subtree B of scale-space atoms, we want to improve our approximation by using one more atom. Again, the question is which atom to choose. We show that the atom must be a child of B . Further, if we have knowledge of $C(d/B)$ for all $d \in \text{chil}(B)$, then we can find the HP estimator conditioned on this knowledge. We show that, given this knowledge, the scale-space atom to choose is the one in $\text{chil}(B)$ that maximizes $C(d/B)$. These results are proved in theorem 4 below.

Theorem 10 (i) Let $F_1 = \{\sum_{m=M+1}^{K-1} c_d \phi_d^m : d \in D_{(0,1]}\}$. Then $\hat{f}_{HP(F_1)} = S_{\{1\}}f$.
(ii) Let B be a subtree with root at 1. Let $g = f - S_B f$. Then $\hat{g}_{HP(F_1)} = S_{\{d\}}g$ where $d \in \text{chil}(B)$.
(iii) Let $d^* = \arg \max_{d \in \text{chil}(B)} \{C_f(d/B)\}$. Then $\hat{g}_{HP(F_1)} = S_{\{d^*\}}g$ where the probability (in the definition of $\hat{g}_{HP(F_1)}$) is now conditioned on knowledge of d^* .

Proof: The proof is in the appendix. ■

Note that in part (iii) of the theorem, the choice of d^* entails maximizing the cost over only the children of a given tree B , not over the complete tree T .

A description of our algorithm is given in figure (4.1). At each step, it works its way down the tree and chooses the atom that maximizes the probability of minimizing the $\mathcal{L}_{(0,1]}^2$ error.

What our algorithm has in common with the conventional multiresolution analysis method is that we first seek information at a coarse level and then zoom in on the details. What distinguishes our method from the multiresolution approach is the following. We recognize the fact that some high resolution coefficients may be more important than other low resolution ones. So we choose coefficients adaptively. Our process of choosing coefficients creates a subtree of scale-space atoms which allows zooming in at one location while having relatively coarse resolution at another location.

6. EXPERIMENTAL RESULTS

In this chapter, we briefly present some experimental results of our work. Further details of our experimental results can be found in [15], [12] [13] [14].

6.1 Compression

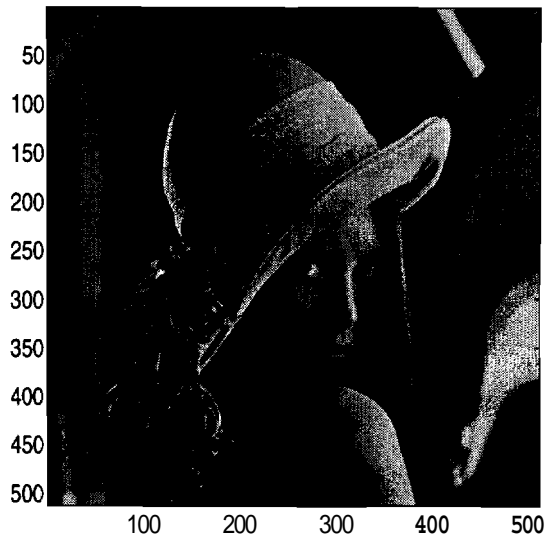
An important property of the multispline compression algorithm is that the representation degrades gracefully as the number of coefficients is decreased. This is shown in figure (6.1).

As can be seen in the figures, if the number of coefficients is reduced from 4000 to 2000, the degradation of image quality is far more acceptable in multispline compression than in wavelet or DCT compression. If the ATM network is being used for a videophone application, the 4000 coefficient multispline representation shown in figure (6.1f) may be adequate. If the DCT or wavelet compression schemes are adopted, the image quality may not be adequate and a larger number of coefficients may need to be transmitted. Now if the network is congested and the ATM switch requests the transmitting user to lower his transmitting rate down the 2000 coefficients per image, the resulting quality will be as shown in figures (6.1 c, e, g). Even if 2000 coefficients are discarded, the multispline compression algorithm still yields a reasonable image, whereas the DCT compressed image has unacceptable quality. An important requirement of the compression algorithm used is graceful degradation. The multispline algorithm is clearly superior in this respect and thus seems to be a better choice than the DCT or wavelet based algorithms.

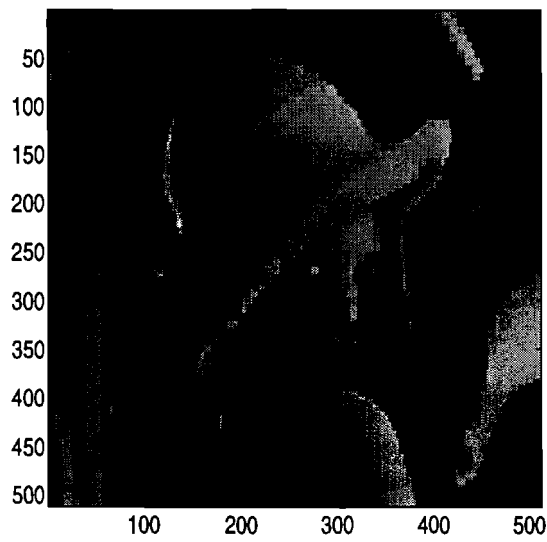
6.2 Noise Removal

The multispline algorithm can also be effectively used to remove noise. In fact, we get compression in addition to noise removal [13].

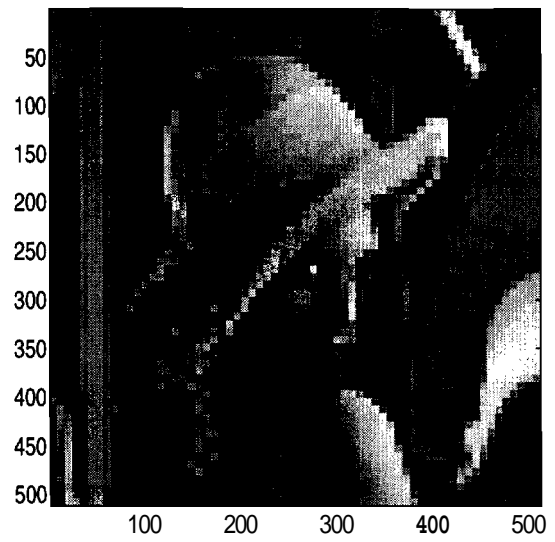
We now present experimental results of simultaneously compressing and de-noising an image. We compare our approach to the restoration technique in [5] since it appears to be the best available algorithm. The approach in [5] is effective on small blocks of the image. However, using the technique this way causes a blocking effect to appear. On the other hand, if we apply the algorithm to the entire image, the blocking disappears but the noise is not removed. This is clearly seen in figures (6.2b,c) which shows the results for Gaussian noise. In Figure (6.2d), we show the results for the multispline algorithm which not only does a better job at de-noising



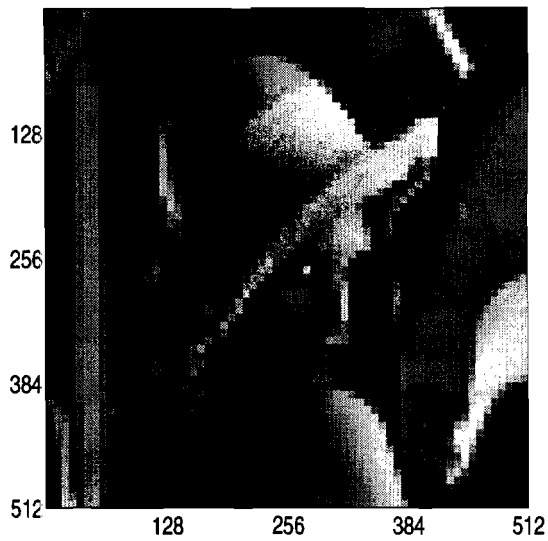
(a) Original Image.



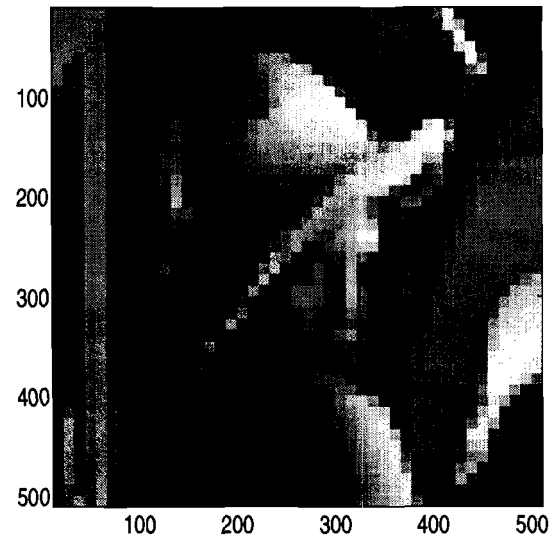
(b) DCT: 4000 coefficients.



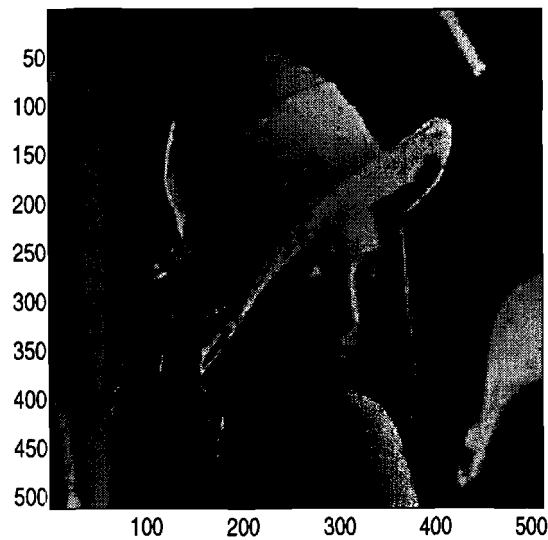
(c) DCT: 2000 coefficients.



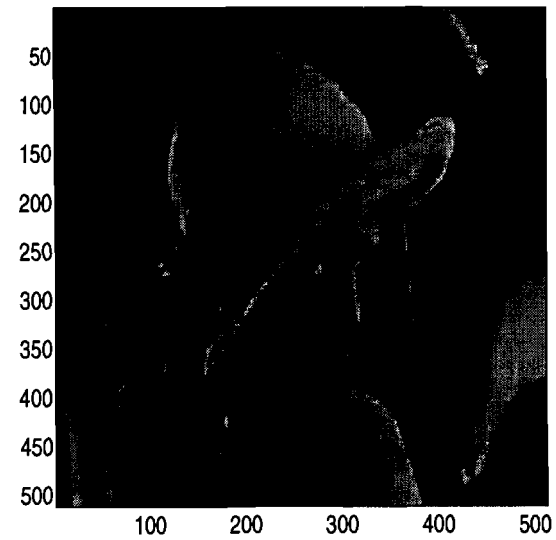
(d) Haar wavelet: 4000 coefficients.



(e) Haar Wavelet: 2000 coefficients.



(f) Multisplines, 4000 coefficients.



(g) Multisplines: 2000 coefficients.

Figure 6.1 Comparison of degradation in image quality as number of coefficients reduced. The Multispline approach outperforms both Haar-wavelet and DCT methods at both high and low compression ratios. Instead of the Haar wavelet, other orthogonal wavelets may be used, but these are also jagged in appearance

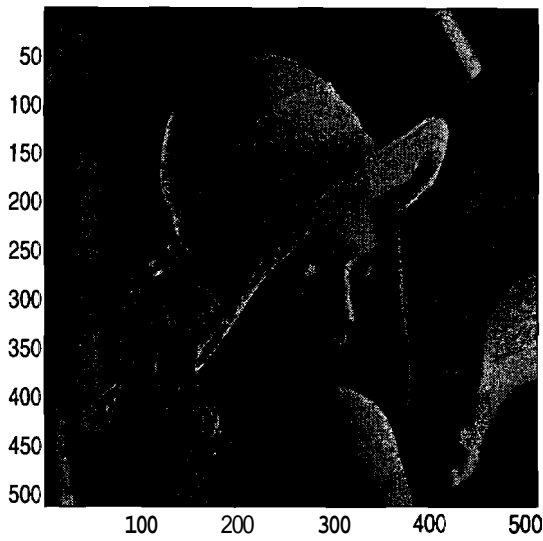
the image but also simultaneously performs image compression. Note that there is no blocking effect in the multispline representation.

6.3 Other Properties

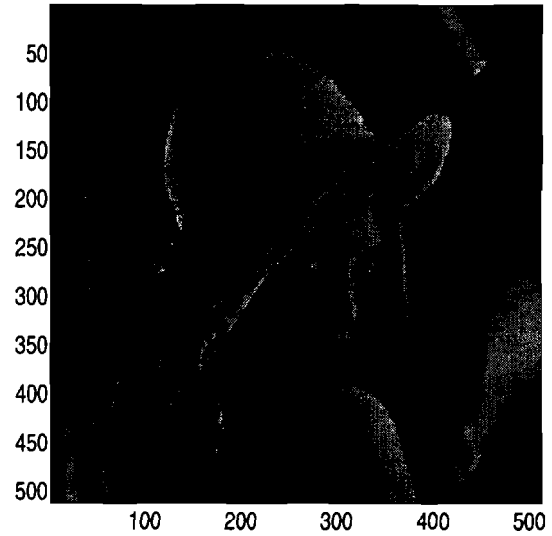
Our experimental results show that multispline compression provides graceful degradation as the number of coefficients is reduced. We also observed that we can simultaneously de-noise and compress an image using multisplines.

Multisplines possess some other properties that make it particularly useful for communication involving multimedia databases. The multispline representation is relatively robust to cell loss. In [13], it is shown that multisplines may be used effectively to combat cell loss over an ATM (Asynchronous Transfer Mode) network. Further, by using multisplines, the overhead associated with coding; the index set is halved (as opposed to the DCT). Another advantage of the multispline representation is that it results in a much lower number of accesses to the hard disk containing the image data. This is particularly useful in browsing applications [12]. Finally, multisplines can also be used to estimate motion in motion image sequences [11].

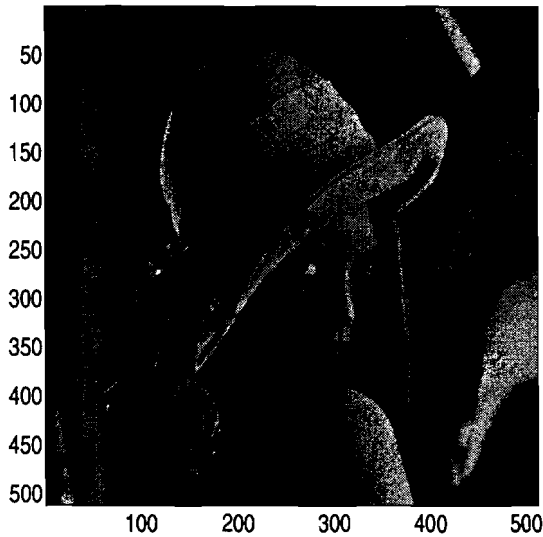
The properties outlined here together with other properties make multisplines particularly useful for image compression and representation.



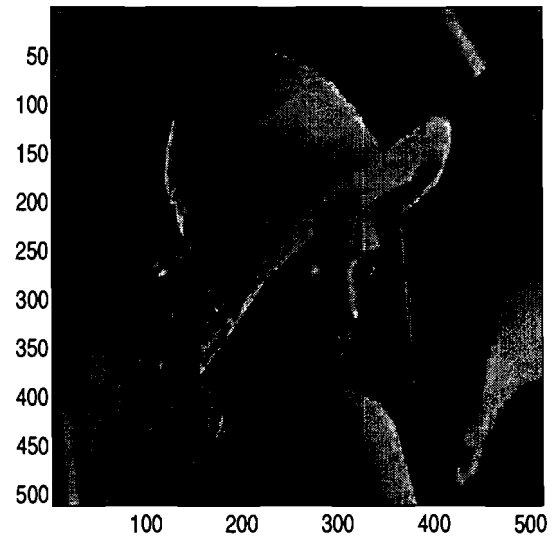
(a) Image with Gaussian noise.



(b) Restoration using Random Field model, 8x8 blocks. No compression.



(c) Restoration using Random Field model on full image. No compression



(d) Joint restoration and compression using Multisplines: 65:1 compression.

Figure 6.2 Gaussian Noise: Comparison of our restoration method to the random field approach. In figure (b), blocking can be observed. In figure (c), there is no blocking but noise is not removed. The multispline approach (figure (d)) has no blocking artifacts, removes noise, and additionally performs compression.

7. CONCLUSIONS

This report developed the theoretical framework for the Ordered-Tree process and its optimal compression algorithm. It has been demonstrated in [12], [14] that this algorithm yields very high compression together with excellent visual quality.

We showed the linear independence of multisplines. We proved that for saturated scale-space sets, multisplines are equivalent to a multiresolution analysis and also to a piecewise polynomial space.

We defined a Tree-Structured analysis and showed that multisplines are an example of a Tree-Structured analysis.

We defined an Ordered-Tree process and proved its existence. We showed an important property that the OTP adds details in any interval infinitely often.

We derived the optimal compression algorithm and explored some of its properties. In particular, it zooms in on edges.

The ordered tree process shows excellent promise in image and signal compression. Future work includes extension of the theoretical and practical results to compressing motion image sequences.

LIST OF REFERENCES



LIST OF REFERENCES

- [1] M. Basseville, A. Benveniste, K. C. Chou, S. A. Golden, R. Niloukhah, and A. S. Willisky. "Modeling and Estimation of Multiresolution Stochastic Processes". *IEEE Transactions on Information Theory*, 38(2):766 – 784, March 1992.
- [2] P. Burt and E. H. Adelson. "The Laplacian Pyramid as a Compact Image Code". *IEEE Transactions on Communications*, 31:532 – 540, 1983.
- [3] C. K. Chui. "An Introduction to Wavelets". Academic Press, San Diego, California, 1992.
- [4] D. L. Donoho. "Nonlinear Wavelet Methods for Recovery of Signals, Densities and Spectra from Indirect and Noisy Data ". In I. Daubechies, editor, *Different Perspectives on Wavelets*, pages 173 – 205. American Mathematical Society, 1993.
- [5] R. L. Kashyap and K. Eom. "Robust Image Models and Their Applications ". In P. W. Hawkes, editor, *Advances in Electronics and Electron Physics*, Vol 70, pages 80 – 157. Academic Press, 1988.
- [6] R. L. Kashyap and S. Moni. "Using the Residual Area Criterion for Signal Reconstruction from Noiseless and Noisy Samples". submitted to *IEEE Transactions on Signal Processing*, 1994.
- [7] B. J. Lucier. "Wavelets and Image Compression ". In T. Lynche and L. L. Schumaker, editors, *Mathematical Methods in CAGD and Image Processing*, pages 1 – 10. Academic Press, 1992.
- [8] S. G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):674 – 693, July 1989.
- [9] S. G. Mallat. "Multiresolution Approximations and Wavelet Orthonormal Bases of $\mathcal{L}^2(\mathbf{R})$ ". *Transactions of the American Mathematical Society*, 315:69 – 87, 1989.
- [10] M. K. Mandal, S. Panchanathan, and T. Aboulnasr. "Wavelets for Image Compression". In *IEEE-SP International Symposium on Time-Frequency Analysis*, pages 338 – 341, Philadelphia, Pennsylvania, October 1994.

- [11] S. Moni and R. L. Kashyap. "Multiresolution Estimation of 3D Motion Parameters ". In IEEE-SP International Symposium on *Time-Frequency Analysis*, pages 310 – 313, Philadelphia, Pennsylvania, October 1994.
- [12] S. Moni and R. L. Kashyap. "Multiresolution Representation Scheme for Multimedia Databases". accepted for publication in ACM Multimedia Systems Journal, 1994.
- [13] S. Moni and R. L. Kashyap. "Image Communication over a Distributed Multimedia System". submitted to IEEE Journal on Selected Areas in *Communications*, May 1995.
- [14] S. Moni and R. L. Kashyap. "Multiresolution using Multisplines for Image Compression ". In IEEE International Conference on Image Processing, Washington, D.C., October 1995.
- [15] S. Moni and R. L. Kashyap. "Multisplines and Adaptive Scale Space Atom Selection for Image Representation". IEEE Signal Processing *Letters*, July 1995.
- [16] G. Strang and V. Strella. "Orthogonal Multiwavelets with Vanishing Moments". In Proceedings of the SPIE Conference on Optical *Engineering*, Orlando, April 1994.
- [17] G. W. Wornell. "Synthesis, Analysis and Processing of Fractal Signals ". PhD thesis, Massachusetts Institute of Technology, August 1991.

APPENDIX

APPENDIX PROOFS OF THEOREMS

In order to prove the theorems, we first need to make some definitions and also state and prove some lemmas. Recall from equation (3.1) that any scale-space atom d may be written as $d = \frac{\kappa(d)}{2^{\gamma(d)}}$, where $\kappa(d)$ is odd. We may enumerate the atoms as follows. For any integer $n \geq 1$, we define

$$\theta_j(n) = \inf\{l \in \mathbf{Z} : 2^l \geq n\} \quad [\text{A.1}]$$

$$\theta_k(n) = \begin{cases} 2(n - 2^{\theta_j(n)-1}) - 1 & \text{if } n > 1 \\ 1 & \text{if } n = 1 \end{cases} \quad [\text{A.2}]$$

Then the n th scale-space atom can be written as

$$d(n) = \frac{\theta_k(n)}{2^{\theta_j(n)}}, \quad n \geq 1 \quad [\text{A.3}]$$

This is an enumeration of all scale-space atoms that does a "raster-scan" of the tree T . Given a node d , its enumeration number is defined as

$$n(d) = \lfloor 2^{\gamma(d)-1} \rfloor + \frac{\kappa(d) + 1}{2} \quad [\text{A.4}]$$

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ will be used to denote the ceiling and floor of a number respectively.

In equation (3.4), the functions were truncated and then normalized so that the resulting functions are unit energy. For the lemmas and theorems that follow, it will be convenient to consider the unnormalized functions. We define

$$\tilde{\phi}_d^m(x) = \phi^m(2^{\gamma(d)}x - \kappa(d) + 1)\chi_{(0,1]}(x)$$

The functions ϕ_d^m and $\tilde{\phi}_d^m$ have the same shape and differ only by a scaling factor. In order to express $\tilde{\phi}_d^m$ in terms of its piecewise polynomial representation, we first write

$$\phi^m(2^{\gamma(d)}x - \kappa(d) + 1) = \sum_{k=\kappa(d)-1}^{m+\kappa(d)-1} \sum_{l=0}^m x^l \Gamma_{l,k}^{(m,d)} \chi_{(\frac{k}{2^{\gamma(d)}}, \frac{k+1}{2^{\gamma(d)}}]} \quad [\text{A.5}]$$

where the $\Gamma_{l,k}^{(m,d)}$ are scalar coefficients. We define $\Gamma_{l,k}^{(m,d)}$ to be zero for values of k outside the range specified in the first sum. The piecewise polynomial representation of $\tilde{\phi}_d^m$ is now obtained by multiplying both sides of equation (A.5) by $\chi_{(0,1]}$.

We will now state and prove some lemmas, which will lead to the proofs of the theorems. The first lemma tells us about the behaviour of the coefficient $\Gamma_{m,k}^{(m,d)}$ of x^m in equation (A.5).

Lemma 1 Let the spline $\tilde{\phi}_d^m$ be expressed as in equation (A.5). Then

$$\begin{aligned} \text{(i)} \quad \Gamma_{m,k}^{(m,d)} &= 2^{m\gamma(d)}(-1)^k \frac{1}{m!} \binom{m-1}{k}, & \kappa(d) - 1 \leq k \leq m + \kappa(d) - 1 \\ \text{(ii)} \quad \Gamma_{m,k}^{(m,d)} &\neq \Gamma_{m,k-1}^{(m,d)}, & \kappa(d) \leq k \leq m + \kappa(d) - 1 \end{aligned}$$

Proof: First consider $d = 1$. Then $\phi_d^m = \phi^m$. It is well known [3] that

$$\phi^m(x) = \frac{x}{m} \phi^{m-1}(x) + \frac{1+m-x}{m} \phi^{m-1}(x-1) \quad [\text{A.6}]$$

We will prove part (i) by induction on m . For $m = 0$, $\phi^0(x) = 1$ on $(0, 1]$. So part (i) holds. Now assume part (i) true upto $m - 1$. From equation (A.6) for ϕ^m , the coefficient of x^m on $(k, k+1]$ is

$$\begin{aligned} \Gamma_{m,k}^{(m,1)} &= \frac{1}{m} \Gamma_{m-1,k}^{(m-1,1)} - \frac{1}{m} \Gamma_{m-1,k-1}^{(m-1,1)} \\ &= \frac{1}{m} \left[\frac{1}{(m-1)!} \binom{m-2}{k} (-1)^k - \frac{1}{(m-1)!} \binom{m-2}{k-1} (-1)^{k-1} \right] \\ &= \frac{1}{m!} \binom{m-1}{k} (-1)^k \end{aligned}$$

This proves part (i) for $d = 1$. For an arbitrary d , we have by equation (A.5) that

$$\begin{aligned} \phi^m(2^{\gamma(d)}x - \kappa(d) + 1) &= \sum_{k=0}^m \sum_{l=0}^m \Gamma_{l,k}^{(m,1)} (2^{\gamma(d)}x - \kappa(d) + 1)^l \chi_{(k,k+1]}(2^{\gamma(d)}x - \kappa(d) + 1) \\ &= \sum_{k=\kappa(d)-1}^{m+\kappa(d)-1} \sum_{l=0}^m \Gamma_{l,k}^{(m,1)} (2^{\gamma(d)}x - \kappa(d) + 1)^l \chi_{(\frac{k}{2^{\gamma(d)}}, \frac{k+1}{2^{\gamma(d)}}]} \end{aligned}$$

The proof of part (i) is complete when we compare the coefficient of x^m in the above equation with that in equation (A.5).

For part (ii), if $m = 0$ then the inequality is trivially satisfied since there is no value of k with $\kappa(d) \leq k \leq m + \kappa(d) - 1$. So let $m \geq 1$ and suppose $\Gamma_{m,k}^{(m,d)} = \Gamma_{m,k-1}^{(m,d)}$. Then

$$2^{m\gamma(d)} \frac{(-1)^k}{m!} \frac{(m-1)!}{(m-1-k)!k!} = 2^{m\gamma(d)} \frac{(-1)^{k-1}}{m!} \frac{(m-1)!}{(m-k)!(k-1)!}$$

Cancelling common terms, we get $k = -(m-k)$, which implies $m = 0$, a contradiction. ■

Lemma 2 Let $\mathbf{J} \geq 0$ be a fixed resolution and let $\Gamma_{l,k}^{(m,d)}$ be as in equation (A.5). Let $\mathbf{H}^{(J,m)}$ be the $2^J \times 2^J$ matrix whose (p, q) th entry is defined by

$$\mathbf{H}^{(J,m)}_{p,q} = \Gamma_{m, \mu_J(p,q)}^{(m,d(p))}$$

where $d(p)$ is as given in equation (A.3) and μ_J is defined as

$$\mu_J(p, q) = \lfloor \frac{q-1}{2^{J-\theta_j(p)}} \rfloor \quad [\text{A.7}]$$

Then $\mathbf{H}^{(J,m)}$ is nonsingular.

Proof: For $m = 0$, the lemma follows from elementary row operations on $\mathbf{H}^{(J,m)}$. So assume $m \geq 1$. We will do the proof by induction on J . Assume $J = 0$. Then $\mathbf{H}^{(0,m)}$ is 1×1 , and

$$\theta_j(1) = 0, \quad \theta_k(1) = 1, \quad d(1) = 1, \quad \mu(1, 1) = 0$$

Since ϕ_d^m is composed of polynomial pieces whose order is exactly m , $\Gamma_{m,0}^{(m,1)} \neq 0$. Therefore $\mathbf{H}^{(0,m)}$ is nonsingular.

Now assume the statement true upto $J - 1$. The first 2^{J-1} rows of $\mathbf{H}^{(J,m)}$ can be formed by writing each column of $\mathbf{H}^{(J-1,m)}$ twice, i.e. for these rows of $\mathbf{H}^{(J,m)}$, the $(2n - 1)$ th and $(2n)$ th columns are identical ($n = 1, \dots, 2^{J-1}$). Therefore,

$$\mathbf{H}^{(J,m)}_{p,2n-1} = \mathbf{H}^{(J,m)}_{p,2n} \quad \forall p, q \in \{1, 2, \dots, 2^{J-1}\} \quad [\text{A.8}]$$

By the induction hypothesis, the first 2^{J-1} rows of $\mathbf{H}^{(J,m)}$ rows are all linearly independent.

For the remaining 2^{J-1} rows of $\mathbf{H}^{(J,m)}$, we have $\theta_j(p) = J$. Therefore, for these rows, we have

$$\mathbf{H}^{(J,m)}_{p,q} = \Gamma_{m,q-1}^{m,d(p)} \quad \forall p \in \{2^{J-1} + 1, \dots, 2^J\}$$

Now consider the $(2^{J-1} + 1)$ th row. For this row, we know (by Lemma 1) that the entries in the first two columns are unequal, i.e.

$$\mathbf{H}^{(J,m)}_{2^{J-1}+1,1} \neq \mathbf{H}^{(J,m)}_{2^{J-1}+1,2} \quad [\text{A.9}]$$

Further, since $\theta_k(p) \geq 3$ for all $p \geq 2^{J-1} + 2$, the entries of the first two columns of $\mathbf{H}^{(J,m)}$ are zero for all these values of p , i.e.

$$\mathbf{H}^{(J,m)}_{p,1} = 0, \quad \mathbf{H}^{(J,m)}_{p,2} = 0, \quad \forall p \geq 2^{J-1} + 2 \quad [\text{A.10}]$$

Further, we have

$$\mathbf{H}^{(J,m)}_{p,1} = \mathbf{H}^{(J,m)}_{p,2} \quad \forall p \leq 2^{J-1} \quad [\text{A.11}]$$

Define vectors \mathbf{v} and $\bar{\mathbf{v}}$ as

$$\mathbf{v} = [v_1, \dots, v_{2^J}], \quad \bar{\mathbf{v}} = [v_1, \dots, v_{2^{J-1}}, \underbrace{0, \dots, 0}_{2^{J-1} \text{ zeros}}]$$

□

Now let

$$\mathbf{v}\mathbf{H}^{(J,m)} = \mathbf{w} = [w_1, \dots, w_{2^J}], \quad \bar{\mathbf{v}}\mathbf{H}^{(J,m)} = \mathbf{u} = [u_1, \dots, u_{2^J}]$$

Note that from equation (A.8) we get

$$u_{2n-1} = u_{2n} \quad \forall n = 1, 2, \dots, 2^{J-1} \quad [\text{A.12}]$$

It now follows from equation (A.10) that

$$w_1 = u_1 + v_{2^{J-1}+1} \mathbf{H}^{(J,m)}_{2^{J-1}+1,1}, \quad w_2 = u_1 + v_{2^{J-1}+1} \mathbf{H}^{(J,m)}_{2^{J-1}+1,2} \quad [\text{A.13}]$$

Suppose $\mathbf{w} = \mathbf{0}$. Then $w_1 = w_2 = 0$ and it follows from equations (A.9, A.13) that

$$v_{2^{J-1}+1} = 0, \quad u_1 = 0$$

Inductively, it follows (using equation (A.12) that

$$v_j = 0 \quad \forall j \in \{2^{J-1} + 1, \dots, 2^J\}, \quad u_j = 0 \quad \forall j \in \{1, \dots, 2^J\}$$

The second equality together with the induction hypothesis imply that $\mathbf{v} = \mathbf{0}$. This completes the proof of the lemma. ■

Theorem 1 Let D be any finite scale-space set. Let ϕ_d^m be as defined in equation (3.4). Then the set of functions $\{\phi_d^m : d \in D, m \in \mathbf{Z}_{M,K}\}$ form a linearly independent set of functions.

Proof: Let

$$f = \sum_{d \in D} \sum_{m \in \mathbf{Z}_{M,K}} c_d^m \tilde{\phi}_d^m \quad [\text{A.14}]$$

and

$$\mathcal{F}_D = \{\tilde{\phi}_d^m : m \in \mathbf{Z}_{M,K}, d \in D\}$$

Let $\mathbf{J} = \max\{\gamma(d) : d \in D\}$ be the resolution of the finest resolution atom in D . By construction, the function f is piecewise polynomial with the degree of each piece at most $K - 1$. Suppose f is identically zero. Then the c_d^m 's must be chosen so that the coefficient of $x^{K-1}, x^{K-2}, \dots, x^{M+1}$ are all zero on each interval $(\frac{q-1}{2^J}, \frac{q}{2^J}]$, $1 \leq q \leq 2^J$.

Now $\{\tilde{\phi}_d^{K-1} : d \in D\}$ are the only elements of \mathcal{F}_D that have a term x^{K-1} . So the c_d^{K-1} 's must be chosen so that the coefficient of x^{K-1} in equation (A.14) is zero. Using equation (A.5), the coefficient of x^{K-1} on the interval $(\frac{q-1}{2^J}, \frac{q}{2^J}]$ is

$$\sum_{d \in D} c_d^{K-1} \Gamma_{K-1, \mu_J(n(d), q)}^{(K-1, d)}$$

where $n(\cdot)$ and $\mu_J(\cdot, \cdot)$ are as defined in equations (A.4) and (A.7) respectively.

By Lemma 2, we know that the matrix $\mathbf{H}^{(J, K-1)}$ is nonsingular. So for the above expression to be zero, we must have $c_d^{K-1} = 0 \quad \forall d \in D$. By backward induction on m , it follows that we have $c_d^m = 0 \quad \forall d \in D, m \in \mathbf{Z}_{M,K}$. Since the ϕ_d^m 's are a scaled version of the $\tilde{\phi}_d^m$'s, the theorem is proved. ■

Lemma 3 Let n be a positive integer. Then

$$\phi^m(2^j x - 2n + 1)\chi_{\{0,1\}} \in \text{span}\{\phi^m(2^{j-1}x - k)\chi_{\{0,1\}}, \phi^m(2^j x - 2l)\chi_{\{0,1\}} : 0 \leq k, l < 2^{j-1}\}$$

Proof: The two-scale relation for splines [3] can be written as

$$\phi^m(2^{j-1}x) = \sum_{\substack{l=0 \\ l \text{ even}}}^{m+1} b^m(l)\phi^m(2^j x - l) + \sum_{\substack{l=0 \\ l \text{ odd}}}^{m+1} b^m(l)\phi^m(2^j x - l) \quad [\text{A.15}]$$

where $b^m(l)$ are the b-spline coefficients which are given by

$$b^m(l) = \begin{cases} 2^{-m} \binom{m+1}{l} & \text{if } 0 \leq l \leq m+1 \\ 0 & \text{otherwise} \end{cases} \quad [\text{A.16}]$$

Let $C_{m,j}^{\text{odd}}$ be a $2^{j-1} \times 2^{j-1}$ matrix whose (p, q) th entry ($1 \leq p, q \leq 2^{j-1}$) is given by

$$(C_{m,j}^{\text{odd}})_{p,q} = \begin{cases} b^m(2(q-p) + 1) & \text{if } 0 \leq q - p \leq 2^{j-1} \\ 0 & \text{otherwise} \end{cases}$$

The entries below the main diagonal are all zero. The entries of the main diagonal are all $(m+1)2^{-m}$. Therefore, the eigenvalues of $C_{m,j}^{\text{odd}}$ are all $(m+1)2^{-m}$ which is strictly positive for all $m \geq 0$. Hence $C_{m,j}^{\text{odd}}$ is invertible.

Define the vectors $\Phi_{m,j-1}$, $\Phi_{m,j}^{\text{even}}$, $\Phi_{m,j}^{\text{odd}}$, \mathbf{e}_n and the $2^{j-1} \times 2^{j-1}$ matrix $C_{m,j}^{\text{even}}$ as follows

$$\begin{aligned} \Phi_{m,j-1}(x) &= [\phi^m(2^{j-1}x)\chi_{\{0,1\}}, \phi^m(2^{j-1}x-1)\chi_{\{0,1\}}, \dots, \phi^m(2^{j-1}x-2^{j-1}+1)\chi_{\{0,1\}}]^T \\ \Phi_{m,j}^{\text{even}}(x) &= [\phi^m(2^j x)\chi_{\{0,1\}}, \phi^m(2^j x-2)\chi_{\{0,1\}}, \dots, \phi^m(2^j x-2^j+2)\chi_{\{0,1\}}]^T \\ \Phi_{m,j}^{\text{odd}}(x) &= [\phi^m(2^j x-1)\chi_{\{0,1\}}, \phi^m(2^j x-3)\chi_{\{0,1\}}, \dots, \phi^m(2^j x-2^j+1)\chi_{\{0,1\}}]^T \\ \mathbf{e}_n &= [0, 0, \dots, 0, \underbrace{1}_{n\text{th spot}}, 0, \dots, 0]^T \end{aligned}$$

$$(C_{m,j}^{\text{even}})_{p,q} = \begin{cases} b^m(2(q-p)) & \text{if } 0 \leq q - p \leq 2^{j-1} \\ 0 & \text{otherwise} \end{cases}$$

Then it follows from equation (A.15) that

$$C_{m,j}^{\text{odd}} \Phi_{m,j}^{\text{odd}} = \Phi_{m,j-1} - C_{m,j}^{\text{even}} \Phi_{m,j}^{\text{even}}$$

If $1 \leq n \leq 2^{j-1}$ then

$$\begin{aligned} \phi^m(2^j x - 2n + 1)\chi_{\{0,1\}} &= \mathbf{e}_n^T \Phi_{m,j}^{\text{odd}}(x) \\ &= \mathbf{e}_n^T [C_{m,j}^{\text{odd}}]^{-1} [\Phi_{m,j-1} - C_{m,j}^{\text{even}} \Phi_{m,j}^{\text{even}}] \end{aligned}$$

which is a linear combination of functions in $\Phi_{m,j-1}$ and $\Phi_{m,j}^{\text{even}}$. If $n > 2^{j-1}$ then

$$\phi^m(2^j x - 2n + 1)\chi_{\{0,1\}} \equiv 0$$

Hence the lemma is proved. ■

Theorem 2 Let D_j be the saturated scale-space set at resolution j . Then

$$\begin{aligned} \text{span}\{\phi_d^m : d \in D_j, m \in \mathbf{Z}_{M,K}\} = \\ \text{span}\{\phi^m(2^j x - k)\chi_{(0,1]} : k = 0, 1, \dots, 2^j - 1, m \in \mathbf{Z}_{M,K}\} \end{aligned}$$

Proof: Using Theorem 1, it is sufficient to prove the statement for a fixed m . We therefore fix m and prove the statement by induction on j . Let

$$\begin{aligned} U_j^m &= \{\phi_d^m : d \in D_j\} \\ V_j^m &= \{\phi^m(2^j x - k)\chi_{(0,1]} : 0 \leq k < 2^j\} \\ V_j^{m,\text{even}} &= \{\phi^m(2^j x - k)\chi_{(0,1]} : k = 0, 2, \dots, 2^j - 2\} \end{aligned}$$

It is clear from the definitions of ϕ_d^m and D_j that

$$\text{span}\{U_j^m\} = \text{span}\{U_{j-1}^m \cup V_j^{m,\text{even}}\}$$

Let $j = 1$. Then

$$\begin{aligned} U_1^m &= \text{span}\{\phi^m(x)\chi_{(0,1]}, \phi^m(2x)\chi_{(0,1]}\} \\ V_1^m &= \text{span}\{\phi^m(2x)\chi_{(0,1]}, \phi^m(2x-1)\chi_{(0,1]}\} \end{aligned}$$

We know that elements of V_j^m are linearly independent. So $\dim\{\text{span}\{V_1^m\}\} = 2$. Since U_1^m contains two functions, $\dim\{\text{span}\{U_1^m\}\} \leq 2$. By lemma 3, $\phi^m(2x-1)\chi_{(0,1]} \in \text{span}\{U_1^m\}$. Therefore, $\text{span}\{V_1^m\} \subset \text{span}\{U_1^m\}$. So $\dim\{\text{span}\{U_1^m\}\} = \dim\{\text{span}\{V_1^m\}\} = 2$ and consequently, $\text{span}\{U_1^m\} = \text{span}\{V_1^m\}$.

Now assume the statement true for $j-1$. Since elements of V_j^m are linearly independent, $\dim\{\text{span}\{V_j^m\}\} = 2^j$. Clearly, $\dim\{\text{span}\{U_j^m\}\} \leq 2j$. By Lemma 3,

$$\begin{aligned} \phi^m(2^j x - k)\chi_{(0,1]} &\in \text{span}\{V_{j-1}^m \cup V_j^{m,\text{even}}\} && \text{for } k = 1, 3, \dots, 2^j - 1 \\ &= \text{span}\{U_{j-1}^m \cup V_j^{m,\text{even}}\} && \text{by induction hypothesis} \\ &= \text{span}\{U_j^m\} && \text{by construction of } U_j^m \end{aligned}$$

By a similar dimensionality argument as in the case of $j = 0$, we get $\text{span}\{U_j^m\} = \text{span}\{V_j^m\}$. ■

Theorem 3 Let M, K be integers with $-1 \leq M < K$. Let D be a saturated scale space set. Then

$$\tilde{\mathcal{S}}_D^{K,M} = \mathcal{P}_D^{K,M}$$

where $\mathcal{P}_D^{K,M}$ is the space of piecewise polynomials functions which are composed of piecewise polynomials of degree at most $K-1$ with knots at the elements of D , and possess M continuous derivatives.

Proof: It follows from Theorem 1 that $\dim\{\mathcal{S}^{0;M}\} = M + 1$ and that $\dim\{\mathcal{S}_D^{K,M}\} = 2^j(K - M - 1)$. It follows (again from Theorem 1) that

$$\dim\{\tilde{\mathcal{S}}_D^{K,M}\} = \dim\{\mathcal{S}^{0;M}\} + \dim\{\mathcal{S}_D^{K,M}\} = M + 1 + 2^j(K - M - 1) \quad [\text{A.17}]$$

The number of derivative constraints on a polynomial in $\mathcal{P}_D^{K,M}$ is $(2^j - 1)(M + 1)$, and these are linear constraints. The number of coefficients for a polynomial of degree $K-1$ is K and there are 2^j polynomial pieces. Therefore

$$\dim\{\mathcal{P}_D^{K,M}\} = 2^j K - (2^j - 1)(M + 1) \quad [\text{A.18}]$$

It is clear that $\mathcal{S}_D^{K,M} \subset \mathcal{P}_D^{K,M}$ because, by construction, $\mathcal{S}_D^{K,M}$ is composed of piecewise polynomials of degree at most $K - 1$ with M continuous derivatives with knots at elements of D . But by equations (A.17, A.18), $\mathcal{S}_D^{K,M}$ and $\mathcal{P}_D^{K,M}$ have the same dimension. Therefore, $\mathcal{S}_D^{K,M} = \mathcal{P}_D^{K,M}$. ■

For the remaining proofs, we will require the Gram matrix of inner products. Let $D = \{d_1, \dots, d_{|D|}\}$ and $\check{D} = \{\check{d}_1, \dots, \check{d}_{|\check{D}|}\}$ be finite scale-space sets. We define the $|D|(K - M + 1) \times |\check{D}|(K - M + 1)$ Grammatrix as

$$(\mathbf{G}_{D,\check{D}})_{d,\check{d}} = \langle \phi_d^m, \phi_{\check{d}}^{\check{m}} \rangle, \quad m \in \mathbf{Z}_{M,K} \quad d \in D, \check{d} \in \check{D}$$

Before stating the next lemma, we introduce some notation. Let D be a scale-space set and let $\mathbf{J} \geq 0$ be some fixed resolution. We define

$$\begin{aligned} D_{1,J} &= \{d \in D : \gamma(d) < J\} \\ D_{2,J} &= \{d \in D : \gamma(d) \geq J\} \end{aligned} \quad [\text{A.19}]$$

A node d is *above* a resolution level \mathbf{J} if $\gamma(d) < J$. A node d is *below* a resolution level \mathbf{J} if $\gamma(d) \geq J$.

We also define some classes of scale-space sets. Let B be a fixed scale-space set (possibly, but not necessarily a subtree). Let $\mathbf{J} \geq 0$ be some fixed resolution. We define

$$\begin{aligned} \mathcal{D}_B &= \{D : |D| = |B|, D \neq B\} \\ \mathcal{D}_{B_{1,J}} &= \{D = \{d_1, d_2, \dots, d_k\} : 1 \leq k \leq B, \gamma(d_i) < J\} \\ \mathcal{D}_{B_{2,J}} &= \{D = \{d_1, d_2, \dots, d_k\} : 1 \leq k \leq B, \gamma(d_i) \geq J\} \end{aligned}$$

Lemma 4 Let B be a finite subtree. Given $\epsilon > 0$, there exists $J_B \in \mathbf{N}$ so that

$$\mathbf{G}_{B,\check{D}}[\mathbf{G}_{\check{D},\check{D}}]^{-1}\mathbf{G}_{\check{D},B} < \epsilon I$$

for all $\check{D} = \{\check{d}_1, \dots, \check{d}_{|\check{D}|}\}$ which satisfy $\gamma(\check{d}_i) \geq J_B \quad \forall i = 1, \dots, |\check{D}|$. The inequality is to be interpreted in the sense of positive definiteness of matrices.

Proof: The support of ϕ_d^m is contained in

$$\left(\frac{\kappa(\check{d}) - C}{2^{\gamma(\check{d})}}, \frac{\kappa(\check{d}) + C}{2^{\gamma(\check{d})}} \right]$$

Let

$$\mathcal{A}_* = \bigcup_{\check{d} \in \check{D}} \left(\frac{\kappa(\check{d}) - C}{2^{\gamma(\check{d})}}, \frac{\kappa(\check{d}) + C}{2^{\gamma(\check{d})}} \right]$$

Let \mathcal{S}_* be the space of functions whose support is contained in \mathcal{A}_* . Let g be any function and let $\mathcal{S}g$ denote its projection onto \mathcal{S}_* . Then $\mathcal{S}g$ is nonzero on a set of (Lebesgue) measure at most

$$2C \sum_{\check{d} \in \check{D}} \frac{1}{2^{\gamma(\check{d})}}$$

It is clear that $\mathcal{S}_{\check{D}} \subset \mathcal{S}_*$. So for any functions g and h we have

$$\begin{aligned} \|\mathcal{S}_{\check{D}}g\|^2 &\leq \|\mathcal{S}_*g\|^2 \\ \langle \mathcal{S}_{\check{D}}g, \mathcal{S}_{\check{D}}h \rangle &\leq \|\mathcal{S}_*g\| \cdot \|\mathcal{S}_*h\| \end{aligned}$$

where the second inequality follows from the first and the Cauchy Schwarz inequality.

Let

$$s = \max_{\check{d} \in \check{B}} \max_{m \in \mathbf{Z}_{M,K}} \sup_{x \in (0,1]} \phi_d^m(x)$$

Since the ϕ_d^m 's are bounded, s is finite. Therefore,

$$\|\mathcal{S}_*\phi_d^m\|^2 \leq 2Cs^2 \sum_{\check{d} \in \check{D}} \frac{1}{2^{\gamma(\check{d})}}$$

Let \mathbf{v} be a vector of length $|B|(M - K + 1)$ whose components are $v_d^m, m \in \mathbf{Z}_{M,K}, d \in B$. Then

$$\begin{aligned} \mathbf{v}^T \mathbf{G}_{B,\check{D}} [\mathbf{G}_{\check{D},\check{D}}]^{-1} \mathbf{G}_{\check{D},B} \mathbf{v} &= \sum_{d,d' \in B} \sum_{m,m' \in \mathbf{Z}_{M,K}} v_d^m \langle \mathcal{S}_{\check{D}}\phi_d^m, \mathcal{S}_{\check{D}}\phi_{d'}^{m'} \rangle v_{d'}^{m'} \\ &\leq \sum_{d,d' \in B} \sum_{m,m' \in \mathbf{Z}_{M,K}} v_d^m \|\mathcal{S}_{\check{D}}\phi_d^m\| \cdot \|\mathcal{S}_{\check{D}}\phi_{d'}^{m'}\| v_{d'}^{m'} \\ &\leq 2Cs^2 \left(\sum_{\check{d} \in \check{D}} \frac{1}{2^{\gamma(\check{d})}} \right) \mathbf{v}^T \mathbf{v} \end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality. Choosing J_B sufficiently large, we can make the quantity in parantheses arbitrarily small. This completes the proof. ■

Lemma 5 Let $N \in \mathbb{N}$ be a fixed number and

$$g(x) = \sum_{d \in B} \sum_{m \in \mathbb{Z}_{M,K}} c_d^m \phi_d^m(x)$$

where B is a finite subtree and the c_d^m 's are (finite) coefficients. Given $\epsilon > 0$ there exists $J_1 \in \mathbb{N}$ so that for all $J \geq J_1$ and for all D such that $|D| = N$,

$$|(\|S_{D_1,J}g\|^2 - \|S_Dg\|^2)| < \epsilon$$

Proof: Let

$$s = \sup_{x \in (0,1]} g(x)$$

Since the c_d^m 's are finite and the ϕ_d^m 's bounded, s is finite. Pick $J_1 \in \mathbb{N}$ so that $2Cs|D| \leq \epsilon 2^{J_1}$. Let $J \geq J_1$. Let

$$\mathcal{A}_* = \bigcup_{d \in D_{2,J}} \left(\frac{\kappa(d) - C}{2^{\gamma(d)}}, \frac{\kappa(d) + C}{2^{\gamma(d)}} \right]$$

Let \mathcal{S}_* be the set of all functions supported on \mathcal{A}_* . Clearly, $D_{2,J} \subset \mathcal{S}_*$. By our choice of J_1 , we have $\|S_*g\|^2 \leq \epsilon$.

Let \oplus denote the direct sum of two spaces and $\dot{\oplus}$ denote the direct sum of two orthogonal spaces. We then have

$$\begin{aligned} \mathcal{S}_D &= \mathcal{S}_{D_1,J} \oplus \mathcal{S}_{D_2,J} \\ &\subset \mathcal{S}_{D_1,J} \oplus \mathcal{S}_* \\ &= [\mathcal{S}_{D_1,J} \cap (\mathcal{S}_*)^c] \dot{\oplus} \mathcal{S}_* \end{aligned}$$

Therefore,

$$\begin{aligned} \|S_Dg\|^2 &= \|S_*g\|^2 + \|S_{D_1,J \cap (\mathcal{S}_*)^c}g\|^2 \\ &\leq \|S_*g\|^2 + \|S_{D_1,J}g\|^2 \end{aligned}$$

Since $\|S_*g\|^2 \leq \epsilon$, we have

$$\|S_{D_1,J}g\|^2 \leq \|S_Dg\|^2 \leq \|S_{D_1,J}g\|^2 + \epsilon$$

which completes the proof. ■

Lemma 6 Let $B = \{d_1, d_2, \dots, d_{|B|}\}$ be a finite subtree and let

$$g(x) = \sum_{d \in B} \sum_{m \in \mathbb{Z}_{M,K}} c_d^m \phi_d^m(x)$$

Then there exists $\delta(B) > 0$ such that

$$\inf_{D \in \mathcal{D}_B} (\|S_Bg\| - \|S_Dg\|) > \delta(B)$$

Proof: If $|B| \leq 1$, the statement is trivial. So let $|B| \geq 2$. We give an iterative method to find $\delta(B)$.

Let $D = \{d_1, d_2, \dots, d_{|B|}\}$. It follows from lemma 3 that we can pick $J_0 \in \mathbb{N}$ so that if $\gamma(d_i) \geq J_0$ for all $d_i \in D$ then

$$\|S_B g\|^2 - \|S_D g\|^2 \geq \frac{1}{2} \|S_B g\|^2$$

This completes the proof for the case where all of the nodes d_i are below resolution level J_0 (i.e. $\gamma(d_i) \geq J_0 \forall d_i \in D$). We set $\delta_0 = \frac{1}{2} \|S_B g\|^2$. This is the zeroth step of the proof.

At the k th step, we will consider the case where k nodes are above resolution J_k and the remaining $|B| - k$ below J_k . We will pick $J_k \geq J_{k-1}$. This will ensure that the choice J_k will also work for all cases previously considered.

At the k th step, define

$$\delta_k^* = \min_{\{\check{d}_1, \check{d}_2, \dots, \check{d}_k\} : \gamma(d_i) \leq J_{k-1}} \|S_B g\|^2 - \|S_{\{\check{d}_1, \check{d}_2, \dots, \check{d}_k\}} g\|^2$$

It follows from property P2 of a Tree-Structured analysis that $\delta_k^* > 0$. It then follows from lemma 4 that we can pick $J_k \geq J_{k-1}$ so that for all $\mathbf{J} \geq J_k$, we have

$$\|S_D g\|^2 - \|S_{D_1, J} g\|^2 \leq \frac{1}{2} \delta_k^* \quad [\text{A.20}]$$

for all $D \in \mathcal{D}_B$.

We now wish to consider the cases where there are at most k nodes above J_k and the remaining nodes below J_k . We only need to consider the cases where there are at most k nodes above J_{k-1} and the remaining nodes below J_k . (This is because the remaining cases have already been considered for previous values of k .)

For this case, we have

$$S_{D_1, J_{k-1}} g = S_{D_1, J_k} g$$

Therefore, by equation (A.20), we have

$$\|S_D g\|^2 \leq \|S_{D_1, J_{k-1}} g\|^2 + \frac{1}{2} \delta_k^*$$

So

$$\begin{aligned} \|S_B g\|^2 - \|S_D g\|^2 &\geq \|S_B g\|^2 - \|S_{D_1, J_{k-1}} g\|^2 - \frac{1}{2} \delta_k^* \\ &\geq \delta_k^* - \frac{1}{2} \delta_k^* \\ &> 0 \end{aligned}$$

Now let

$$\delta_k = \min\{\delta_0, \delta_1, \dots, \delta_{k-1}, \delta_k^*\}$$

For any $D \in \mathcal{D}_B$ which has at most k nodes above J_k , we have

$$\|S_B g\|^2 - \|S_D g\|^2 \geq \delta_k > 0$$

After the $|B|$ th step, we set $\delta(B) = \delta_{|B|}$ and this completes the proof. ■

Lemma 7 Let B be a finite subtree. Let B^* also be some finite subtree and d^* a child of B^* . Let c be a real number and define

$$\begin{aligned} g(x) &= \sum_{d \in B^*} \sum_{m \in \mathbf{Z}_{M,K}} \phi_d^m(x) \\ h(x) &= \sum_{m \in \mathbf{Z}_{M,K}} c \phi_{d^*}^m(x) \\ \zeta_{B,g}(c) &= \inf_{D \in \mathcal{D}_B} \|S_D(g+h)\|^2 - \|S_D g\|^2 \end{aligned}$$

Then $\zeta_{B,g}(\cdot)$ is continuous at 0.

Proof: Let $D, \check{D} \in \mathcal{D}_B$. Fix $\epsilon > 0$. Pick $J \in \mathbf{N}$ so that

$$\begin{aligned} \|S_{D_{1,J}}(g+h)\|^2 &\leq \|S_D(g+h)\|^2 \leq \|S_{D_{1,J}}(g+h)\|^2 + \epsilon \\ \|S_{\check{D}_{1,J}}g\|^2 &\leq \|S_{\check{D}}g\|^2 \leq \|S_{\check{D}_{1,J}}g\|^2 + \epsilon \end{aligned}$$

for all $D, \check{D} \in \mathcal{D}_B$. We then have

$$\begin{aligned} \inf_{D \in \mathcal{D}_B} \|S_D(g+h)\|^2 &\leq \inf_{D \in \mathcal{D}_B} \|S_{D_{1,J}}(g+h)\|^2 + 2\epsilon \\ \inf_{\check{D} \in \mathcal{D}_B} \|S_{\check{D}_{1,J}}g\|^2 &\leq \inf_{\check{D} \in \mathcal{D}_B} \|S_{\check{D}}g\|^2 \end{aligned}$$

Therefore

$$\inf_{D \in \mathcal{D}_B} \|S_D(g+h)\|^2 - \inf_{\check{D} \in \mathcal{D}_B} \|S_{\check{D}}g\|^2 \leq \inf_{D \in \mathcal{D}_B} \|S_{D_{1,J}}(g+h)\|^2 - \inf_{\check{D} \in \mathcal{D}_B} \|S_{\check{D}_{1,J}}g\|^2 + 2\epsilon$$

There are only a finite number of $D_{1,J}, \check{D}_{1,J}$ in \mathcal{D}_B . Further, $\|S_{D_{1,J}}(g+h)\|^2$ is a continuous function of c for any fixed D . Therefore,

$$\inf_{D \in \mathcal{D}_B} \|S_{D_{1,J}}(g+h)\|^2 - \inf_{\check{D} \in \mathcal{D}_B} \|S_{\check{D}_{1,J}}g\|^2$$

is a continuous function of c and is zero at $c = 0$. We can therefore pick $\delta_1 > 0$ so that if $|c| < \delta_1$ then

$$\left| \inf_{D \in \mathcal{D}_B} \|S_{D_{1,J}}(g+h)\|^2 - \inf_{\check{D} \in \mathcal{D}_B} \|S_{\check{D}_{1,J}}g\|^2 \right| < \epsilon$$

For such a c , we have by equation (A.21),

$$\left| \inf_{D \in \mathcal{D}_B} \|S_D(g+h)\|^2 - \inf_{\check{D} \in \mathcal{D}_B} \|S_{\check{D}}g\|^2 \right| < 3\epsilon$$

Now

$$\|S_B(g+h)\|^2 - \|S_B g\|^2$$

is a continuous function of \mathbf{c} and zero at $\mathbf{c} = \mathbf{0}$. So we can pick $\delta_2 > 0$ so that if $|c| < \delta_2$ then

$$\left| \|S_B(g+h)\|^2 - \|S_B g\|^2 \right| < \epsilon$$

If $|c| \leq \min(\delta_1, \delta_2)$ then we have

$$\begin{aligned} |\zeta_{B,g}(c) - \zeta_{B,g}(c)| &= \left| \|S_B(g+h)\|^2 - \inf_{D \in \mathcal{D}_B} \|S_D(g+h)\|^2 - \|S_B g\|^2 + \inf_{D \in \mathcal{D}_B} \|S_D g\|^2 \right| \\ &\leq \left| \|S_B(g+h)\|^2 - \|S_B g\|^2 \right| + \left| \inf_{D \in \mathcal{D}_B} \|S_D(g+h)\|^2 - \inf_{D \in \mathcal{D}_B} \|S_D g\|^2 \right| \\ &\leq \epsilon + 3\epsilon = 4\epsilon. \end{aligned}$$

Therefore, $\zeta_{B,g}(\cdot)$ is continuous at 0. ■

Theorem 4 There exists an Ordered-Tree process $g(x)$ as in equation (4.3) satisfying equation (4.4).

Proof: We first construct a random ordering of the nodes of the tree T with the property that for any k , the first k nodes form a subtree. At the first step, We get a subtree B_1 by picking the node $d = 1$. At the n th stage, we get a subtree B_n by picking one of the nodes in $\text{chil}(B_{n-1})$. At the n th stage, the choice of the node is random and satisfies the condition that all nodes in $\text{chil}(B_{n-1})$ have equal probability of getting picked. It is easy to see that this procedure imposes an ordering on the nodes of the tree T with the property that for any k , the first k nodes form a subtree.

Let d_i be the i th node picked by the random ordering described above. We construct the ordered tree process by constructing a sequence of functions

$$g_n(x) = \sum_{i=1}^n \sum_{m \in \mathbf{Z}_{M,K}} c_{d_i}^m \phi_{d_i}^m(x)$$

and taking the limit as n goes to infinity. The function g is formed by picking the coefficients $c_{d_n}^m$ for the functions $\phi_{d_n}^m$. Note that we have to pick $K - M + 1$ coefficients, one for each value of m in $\mathbf{Z}_{M,K}$.

In this proof, we will set the coefficients $c_{d_n}^m$ to be the same for all $m \in \mathbf{Z}_{M,K}$. That is, the coefficient will depend upon the node d , but not on the order m of the spline. We will require the sequences $\{w_n\}$ and $\{\alpha_n\}$ which we define as

$$\begin{aligned} w_n &= \frac{1}{2} \left(1 + \frac{1}{n}\right), & n \geq 1 \\ \alpha_n &= \frac{1}{2^{n/2}}, & n \geq 1 \end{aligned}$$

Note that $w_1 = 1$ and $w_n \downarrow \frac{1}{2}$ as $n \rightarrow \infty$.

T